Ensemble streamflow forecasting over a cascade reservoir catchment with integrated hydrometeorological modeling and machine learning

Junjiang Liu[1], Xing Yuan[1,2*], Junhan Zeng[1], Yang Jiao[1], Yong Li[3], Lihua Zhong[3], Ling Yao[4]

[1]School of Hydrology and Water Resources, Nanjing University of Information Science and Technology, Nanjing 210044, China

[2]Key Laboratory of Regional Climate-Environment for Temperate East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

[3]Guangxi Meteorological Disaster Prevention Center, Nanning 530022, China

[4]Guangxi Guiguan Electric Power Co., Ltd., Nanning 530029, China

Hydrology and Earth System Sciences

October 12, 2021

**Corresponding author address:* Xing Yuan, School of Hydrology and Water Resources, Nanjing University of Information Science and Technology, Nanjing 210044, China E-mail: xyuan@nuist.edu.cn

Abstract. A popular way to forecast streamflow is to use bias-corrected meteorological forecast to drive a calibrated hydrological model, but these hydrometeorological approaches have deficiency over small catchments due to uncertainty in meteorological forecasts and errors from hydrological models, especially over catchments that are regulated by dams and reservoirs. For a cascade reservoir catchment, the discharge of the upstream reservoir contributes to an important part of the streamflow over the downstream areas, which makes it tremendously hard to explore the added value of meteorological forecasts. Here, we integrate the meteorological forecast, land surface hydrological model simulation and machine learning to forecast hourly streamflow over the Yantan catchment, where the streamflow is influenced both by the upstream reservoir water release and the rainfall-runoff processes within the catchment. Evaluation of the hourly streamflow hindcasts during the rainy seasons of 2013-2017 shows that the hydrometeorological ensemble forecast approach reduces probabilistic and deterministic forecast errors by 6% as compared with the traditional ensemble streamflow prediction (ESP) approach during the first 7 days. The deterministic forecast error can be further reduced by 6% in the first 72 hours when combining the hydrometeorological forecast with the long short-term memory (LSTM) deep learning method. However, the forecast skill for LSTM using only historical observations drops sharply after the first 24 hours. This study implies the potential of improving flood forecast over a cascade reservoir catchment by integrating meteorological forecast, hydrological modeling and machine learning.

Keywords: Streamflow; Hydrological modeling; LSTM; Reservoir; Ensemble forecast

1. Introduction

Flood events are the most destructive ones among the natural disasters, causing huge damages to human society. Reservoirs are massively constructed to regulate river flows, which has significantly reduced flood risks or damages (Ji et al., 2020). However, the number and intensity of precipitation extreme events are increasing in many areas as the global warming continues, thus amplify the potential of flood hazards (Hao et al., 2013; Shao et al., 2016; Wei et al., 2018; Yuan et al., 2018a; Wang et al., 2019). Accurate streamflow forecast is thus needed to provide guidelines for reservoir operations (Robertson et al., 2013), especially when the flood risk is increasing under global warming.

A common approach of streamflow forecast is to use hydrological models, where the first attempt could be traced back to 1850s, using simple regression-type approaches to predict discharge from observed precipitation (Mulvaney, 1850). Since then, model concepts have been further augmented by designing new data networks, addressing heterogeneity of hydrological processes, capturing the nonlinear characteristics of hydrologic system and parameterizing models (Hornberger and Boyer, 1995; Kirchner, 2006). With the advancements of computer technology and high-resolution observation, a well-parameterized hydrological model can simulate streamflow with high accuracy (Kollet et al., 2010; Ye et al., 2014; Graaf et al., 2015; Yuan et al., 2018b).

Streamflow simulations from hydrological models heavily rely on meteorological forcing inputs, especially precipitation, which can be measured at in-situ gauges or retrieved from satellites and radars. However, for medium-range (2–15 days ahead) streamflow forecasts, precipitation forecast is needed (Hopson et al., 2002). To improve the forecast, ensemble techniques that can give a deterministic estimate as well as its uncertainty became popular. Ensemble weather forecasting can be traced back to 1963 when Leith transferred a deterministic forecast into an ensemble using the Monte-Carlo method to describe the atmospheric uncertainty (Leith, 1963). In the 1990s, ensemble forecasting was developed into an integral part of numerical weather prediction, which showed higher skill than the deterministic forecast even with higher model resolution (Toth et al., 2001). Due to its rapid development, ensemble weather forecasts and climate predictions are applied to hydrological forecasting studies by combining with hydrological models (Jasper et al., 2002; Balint et al., 2006; Jaun et al., 2008; Xu et al., 2015; Yuan et al., 2016; Zhu et al., 2019). Provided with streamflow variability, a reservoir can maintain a reliable utility from natural streamflow better than provided with a deterministic streamflow forecast (Zhao et al., 2011). However, the streamflow prediction skill depends on whether the precipitation forecasts introduced into the hydrological model are skillful (Alfieri et al., 2013). When assessing the skill of this hydrometeorological forecast approach, a benchmark is needed. Using ensembles of historical climatology data (Day, 1985) as meteorological forecast inputs, which is known as ensemble streamflow prediction (ESP), is often selected as the benchmark approach.

Evaluations of hydrological forecasts indicated that forecast skill has a close relationship with catchment size, geographical locations and resolutions (Alfieri et al., 2013; Pappenberger et al., 2015), which means there is a necessity to compare with the ESP to show the skill of the hydrometeorological forecast approach.

Although physically based hydrological models are widely used, it is still hard to apply a hyper-resolution distributed model for streamflow forecasting due to its demand for observation data, complex model structures and computational resources requirements for calibration and application (Wood et al., 2011; Kratzert et al., 2018; Yaseen et al., 2018). In cascade reservoir systems, there are two sources of streamflow, one is from the rainfall within the interval basin and the other is from the upstream reservoir discharge. While the rainfall-runoff relationship is well studied, it is challenging to reproduce the reservoir operating rules in a physical model (Gao et al., 2010; Zhang et al., 2016; Dang et al., 2020).

Machine learning methods can recognize patterns hidden in input data and can simulate or predict streamflow without explicit descriptions of the underlying physical processes (Kisi et al., 2007; Adnan et al., 2019). Neural networks are suitable for streamflow forecasting among machine learning models, some of them can even outperform physically based hydrological models. For example, Humphrey et al. (2016) showed that their combined Bayesian artificial neural network with the modèle du Génie Rural à 4 paramètres Journalier (GR4J) approach outperforms the GR4J model in monthly streamflow forecasting. Kratzert et al. (2019) showed that the long short-term memory (LSTM)-based approach outperforms a well-calibrated Sacramento Soil Moisture Accounting Model (SAC-SMA). Yang et al. (2020) used the geomorphology-based hydrological model (GBHM) combined with traditional ANN model to simulate daily streamflow, which can provide enough physical evidence and can run with less observation data. Although neural network models are criticized with little physical evidence (Abrahart et al., 2012), their potential in hydrological forecasting is yet to be explored.

In this study, we combine the machine learning with hydrometeorological approach for hourly streamflow forecast over a cascade reservoir catchment located in southwestern China. We use the meteorological hindcast data from European Centre for Medium-Range Weather Forecasts (ECMWF) model that participated in the THORPEX Interactive Grand Global Ensemble (TIGGE) project to drive a newly developed high-resolution land surface model, named as the Conjunctive Surface-Subsurface Process model version 2 (CSSPv2, Yuan et al., 2018b), to provide runoff and streamflow forecasts, and correct the forecasts via LSTM model. We aim to improving flood forecast over the cascade reservoir catchment by integrating meteorological forecast, hydrological modeling and machine learning. So we strive to (1) calibrate the hydrological model, (2) bias correct the meteorological forecasts, (3) evaluate the streamflow forecast skill and (4) test the physical-statistical combined approach.

2. Study Area, Data, Model and Method

2.1 Study Area

The Yantan Hydropower Station is in the middle reaches of Hongshui River in Dahua Yao Autonomous County, Guangxi Province. The Yantan Hydropower Station is the fifth level in the 10-level development of Hongshuihe hydropower base in Nanpanjiang River, connected with upstream Longtan Hydropower Station and the downstream Dahua Hydropower Station. The drainage area between the Longtan Hydropower Station and Yantan Hydropower Station is 8,900 km^2. The annual mean streamflow at Yantan gauge is 55.5 billion m^3. The river passes through karst mountain area, with narrow valley, steep slope and scattered cultivated land, and the average slope is 0.036%. Figure 1 shows the locations of 4 hydrological gauges, with detailed information listed in Table 1.

2.2 Data and Method

2.2.1 Hydrometeorological observations

There are 97 meteorological observation stations within the catchment (Figure 1). Here, observed hourly 2m-temperature, 10m-wind speed, relative humidity, accumulated precipitation and surface pressure data were interpolated into a 5km gridded observation dataset via inverse distance weight method. The hourly surface downward solar radiation data from China Meteorological Administration Land Data Assimilation System (CLDAS) was also interpolated into 5km via bilinear interpolation method. The hourly surface downward thermal radiation (long) was estimated by specific humidity, pressure, temperature. This dataset was used to drive the CSSPv2 land surface hydrological model.

The monthly runoff for each 5km grid was estimated by disaggregating control streamflow station observations with the ratio of observed grid monthly precipitation and catchment mean precipitation. The gridded runoff was used to calibrate the CSSPv2 model at each grid (Yuan et al., 2016), which would generate distributed model parameters that are different within the catchment to better represent the heterogeneity of the rainfall-runoff processes.

2.2.2 Ensemble Meteorological hindcast data and ESP hindcasts

The TIGGE dataset consists of ensemble forecast data from 10 global Numerical Weather Prediction centers started from October 2006, which has been made available for scientific research, via data archive portals at ECMWF and the Chine Meteorological Administration (CMA). TIGGE has become a focal point for a range of research projects, including research on ensemble forecasting, predictability, and the development of products to improve the prediction of severe weather (Bougeault et al., 2010). In this paper, TIGGE data from April to September during 2013-2017 from ECMWF were used as meteorological hindcast data. The 3-hourly meteorological hindcasts for 7-day lead time from 51 ensemble members (including control forecast) were interpolated into 5km resolution via bilinear interpolation. The forecast precipitation and temperature were corrected to match the observational means to remove the biases.

The ESP was accomplished by applying historical meteorological forcings (Day, 1985). In this paper, the meteorological forcings from the same date as the forecast start date to the next 9 days of each year (excluding the target year) were selected as the ESP forcings. Take April 1^{st}, 2013 as example, the 7-day observations started from April 1^{st} to April 10^{th} (i.e., April 1^{st}-April 7^{th}, April 2^{nd}-April 8^{th}, …, April 10^{th}-April 16^{th}) in the year of 2014, 2015, 2016 and 2017 were selected as the forecast ensemble forcings of the issue date (April 1^{st}), with a total of 40 ensemble members. The detailed information about the raw datasets are listed in Table 2

2.2.3 CSSPv2 streamflow hindcasts

The physical hydrological model used in this paper is the Conjunctive Surface-Subsurface Process model version 2 (CSSPv2; Yuan et al., 2018). The CSSPv2 model is a distributed, grid-based land surface hydrological model, which was developed from the Common Land Model (Dai et al., 2003, 2004), but with better representations in lateral surface and subsurface hydrological processes and their interactions. The routing model used here employs the kinetic wave equation as covariance function, which is solved via a Newton algorithm. A main reason for adopting this covariance function is that it suits the basin with mountainous terrain. The CSSPv2 model was successfully used to perform a high-resolution (3 km) land surface simulation over the Sanjiangyuan region, which is the headwater of major Chinese rivers (Ji and Yuan, 2018). In this paper, we calibrated CSSPv2 model against monthly estimated runoff to simulate the natural hydrological processes by using the

Shuffled Complex Evolution (SCE-UA) approach (Duan et al., 1994). The calibrated parameters include maximum velocity of baseflow, variable infiltration curve parameter, fraction of maximum soil moisture where non-linear baseflow occurs and fraction of maximum velocity of baseflow where non-linear baseflow begins. The hourly observed streamflow at Yantan hydrological gauge was used to calibrate the CSSPv2 routing model manually, including slope, river density, roughness, width and depth. The observed streamflow at Longtan hydrological gauge were added into the corresponding grid to provide upstream streamflow information. We used a high-resolution elevation database (hereafter referred to as DEM30) for sub-grid parameterization and figured out the initial values of these river channel parameters. We first extracted the slope angle and the natural river flow path from DEM30, and then identified the accurate river network using a drainage area threshold of 0.18 km^2. River density and bed slope values for each 5km grid were calculated as:

$$\boldsymbol{rivden} = \sum \boldsymbol{l}/\boldsymbol{A}, \tag{1}$$

$$\boldsymbol{bedslp} = \boldsymbol{mean}(\mathbf{tan}(\boldsymbol{\beta})), \tag{2}$$

where rivden is the river density (km/km^2), bedslp is the river channel bed slope (unitless), A is the area of a 5km grid (km^2), $\sum \boldsymbol{l}$ is the total river channel length (m) within the grid, β is the slope angle (radian) for each river segment located in the grid.

Other river channel parameters were estimated by empirical formulas (Getirana et al., 2012; Luo et al., 2017) as follows:

$$\boldsymbol{W} = \mathbf{1.956} \times \boldsymbol{A_{acc}^{0.413}}, \tag{3}$$

$$\boldsymbol{H} = \mathbf{0.245} \times \boldsymbol{A_{acc}^{0.342}}, \tag{4}$$

$$\boldsymbol{n} = \mathbf{0.03} + (\mathbf{0.05} - \mathbf{0.03})\frac{\boldsymbol{H_{max}} - \boldsymbol{H}}{\boldsymbol{H_{max}} - \boldsymbol{H_{min}}}, \tag{5}$$

where W, H and n are river width (m), depth (m) and roughness (unitless) for each 5km grid; Aacc means the upstream drainage area (km2); Hmax and Hmin refer to the maximum and minimum values of river depth calculated by Eq. (4).

Through a trial-and-error procedure, we calibrated these river channel parameters to match the simulated streamflow with observed hourly records at Yantan hydrological gauge. The simulation results were evaluated by calculating the Nash-Sutcliffe efficiency (NSE) with corresponding observation data. The descriptions of the calibrated parameters and their range are listed in Table 3

After calibration, we drove the CSSPv2 model using 5km regridded and bias-corrected TIGGE-ECMWF forecast forcing during 2013-2017 to provide a set of 7-day hindcasts. Streamflow hindcasts both from the ESP and the hydrometerological approach (TIGGE-ECMWF/CSSPv2) were corrected by matching monthly mean streamflow observations to remove the biases, and the hindcast experiments were termed as ESP-Hydro and Meteo-Hydro (Table 4). Figure 2 shows the procession of the CSSPv2 hindcasts: the calibrated CSSPv2 model was first driven with observation dataset to generate initial hydrological conditions (soil moisture, surface water, etc.) for each forecast issue date, then CSSPv2 model was driven with forecast data

(TIGGE-ECMWF or ESP) at every forecast issue date with the generated initial conditions to perform a 7-day hindcast.

2.2.4 LSTM streamflow forecast

LSTM is a type of recurrent neural network model which learns from sequential data. The input of the LSTM model includes forecast interval streamflow at the specified forecast step obtained from TIGGE-ECMWF/CSSPv2, historical upstream streamflow observation, and historical streamflow observation at Yantan hydrological gauge. The network was trained on sequences of April to September in 2013-2017, with six historical streamflow observations and one forecast interval streamflow to predict the total streamflow at each forecast time step (Figure 2). The LSTM was calibrated through a cross validation method, by leaving the target year out.

Before calibration, all input and output variables were normalized as follows:

$$\mathbf{q_0} = \frac{(\mathbf{q}-\mathbf{q_{min}})}{(\mathbf{q_{max}}-\mathbf{q_{min}})}, \quad (6)$$

Where $\mathbf{q_0}$, $\mathbf{q}$, $\mathbf{q_{max}}$ and $\mathbf{q_{min}}$ are the normalized variable, input variable, the maximum and minimum of the sequence of the variable. The hindcast experiment was termed as Meteo-Hydro-LSTM (Table 2). In addition, we also tried an LSTM streamflow forecast approach which only uses 6-hr historical streamflow data as inputs, and the experiment was termed as LSTM (Table 2). The process of LSTM is similar to Meteo-Hydro-LSTM but without the forecast interval streamflow, which is also shown in Figure 2.

2.3 Evaluation Method

The root-mean squared error (RMSE) was used to evaluate the deterministic forecast, i.e., the ensemble means of 51 (ECMWF) or 40 (ESP) forecast members. To evaluate probabilistic forecasts, the Continuous Ranked Probability Score (CRPS) was calculated as follows:

$$\boldsymbol{CRPS} = \int_{-\infty}^{\infty} [\boldsymbol{F}(\boldsymbol{y}) - \boldsymbol{F_o}(\boldsymbol{y})]^{\mathbf{2}}, \quad (7)$$

where

$$\boldsymbol{F_o}(\boldsymbol{y}) = \begin{cases} \mathbf{0}, & \boldsymbol{y} < \boldsymbol{observed\ value} \\ \mathbf{1}, & \boldsymbol{y} \geq \boldsymbol{observed\ value} \end{cases} \quad (8)$$

is a cumulative-probability step function that jumps from 0 to 1 at the point where the forecast variable $\boldsymbol{y}$ equals the observation and $\boldsymbol{F}(\boldsymbol{y})$ is a cumulative-probability distribution curve formed by the forecast ensembles. The CRPS has a negative orientation (smaller values are better), and it rewards concentration of probability around the step function located at the observed value (Wilks, 2005). The skill score for deterministic forecast was calculated as

$$\boldsymbol{SS_{RMSE}} = \frac{\boldsymbol{RMSE} - \boldsymbol{RMSE_{ref}}}{\mathbf{0} - \boldsymbol{RMSE_{ref}}} = \mathbf{1} - \frac{\boldsymbol{RMSE}}{\boldsymbol{RMSE_{ref}}} \quad . \quad (9)$$

The skill score for probabilistic forecast (CRPSS) could be calculated similarly based the CRPS.

3. Results

3.1 Evaluation of CSSP calibration

The employed CSSPv2 model is a fully distributed hydrological model and the streamflow is calculated through a process of converting gridded rainfall into runoff and a process of runoff routing. Figure 3 shows the runoff calibration results by calculating the NSE of monthly runoff simulations compared with observed gridded monthly runoff. After calibrating the CSSPv2 runoff model, the NSE of all grids are above 0, which indicates that the runoff simulation results in all grids are more reliable than the climatology method. In addition, grids distributed in the downstream region have better NSE than the upstream grids. The NSE values of the grids in the southern part are greater than 0.5, which accounts for two thirds of the interval basin area. Higher NSE in the upstream part of Jiazhuan station (Figure 1) is due to more humid climate (not shown), where hydrological models usually have better performance over wetter areas. For the downstream areas with less precipitation, the higher NSE is related to the higher percentage of sand in the soil (not shown). Under the same meteorological conditions, there is higher hydraulic conductivity with higher sand content (Wang et al., 2016), and it yields less runoff under infiltration excess, which is more suitable for the saturation excess-based runoff generation for the CSSPv2 model (Yuan et al., 2018b).

Figures 4 and 5 show the results after the calibration of the routing model, where CSSPv2 is driven by observed meteorological forcings to provide streamflow simulations and compare against observed streamflow at Yantan hydrological gauge. Figure 4 shows the daily and monthly streamflow simulation results. The monthly result (Fig. 4f) shows that the simulated streamflow closely follows the observed streamflow, and the NSE is 0.96. The daily streamflow simulations during flood seasons (Figs. 4a-4e) also show a good performance, and the NSE is 0.92. During June and July in years of 2014, 2015 and 2017, the CSSPv2 model underestimated the daily streamflow with a maximum of 1104 m^3/s and an average of 334 m^3/s (Figs. 4b, 4c, 4e). In years of 2013 and 2016, the difference between observed and simulated streamflow is relatively small, and the average difference is 96 m^3/s (Figs. 4a, 4d).

Figure 5 shows the hourly streamflow simulation results for a few flood events. Figure 5a shows that the CSSPv2 model can accurately simulate the streamflow response to a rainfall event after a dry period. Figures 5b-5d show that for instantaneous heavy rainfall events, the CSSPv2 model over-predicted the water loss during recession period. Figures 5e-5f show that for continuous rainfall events, the simulated streamflow has a larger fluctuation than observation. The simulated streamflow is also smoother than observation. Nevertheless, the NSE for the hourly streamflow simulation is 0.61, which suggests that CSSPv2 has an acceptable performance at hourly time scale.

3.2 Bias correction of TIGGE-ECMWF meteorological forecasts

The resolution of TIGGE-ECMWF grid data is 0.25°, so the data was interpolated to 5km grid to drive the CSSPv2 model. We calculated both observations' and TIGGE-ECMWF's yearly average precipitation and temperature, then performed a bias correction by adding back the difference (for temperature) or multiplying back the ratio (for temperature) to match the observations' averages. Figure 6 shows the correlation coefficient and RMSE of TIGGE-ECMWF precipitation and temperature forecasts as compared against observations, either before or after bias correction. The 51-ensemble mean precipitation and temperature (the red dashed lines) shows better performance than the best ensemble members (the green dashed lines), with an average RMSE reduction of 3.66 mm/day and average correlation increase of 0.04 for precipitation, and average RMSE reduction of 0.1K and average correlation increase of 0.03 for temperature. After bias correction, the 51-ensemble means still perform better than best ensemble members. Compared with ensemble mean results before bias correction, the RMSE reduced by 0.23 mm/day for the bias-corrected precipitation, and reduced by 1K for the bias-corrected surface air temperature. For the bias-corrected ensemble mean results, the average RMSE and correlation are 14.6 mm/day and 0.44 for precipitation, and 1.25 K and 0.87 for surface air temperature.

3.3 Comparison between ESP-Hydro and Meteo-Hydro streamflow forecast

Figure 7 presents the variations of RMSE and CRPS for ESP-Hydro and Meteo-Hydro hourly streamflow forecast at Yantan hydrological gauge. For probabilistic forecast, Figure 7a shows that the CRPS for Meteo-Hydro streamflow forecast ranges from 165 to 225 m^3/s while the CRPS for ESP-Hydro streamflow forecast ranges from 170 to 230 m^3/s. The Meteo-Hydro approach performs better than ESP-Hydro with lower CRPS at all lead times, with an average of 6% improvement in CRPSS (Figure 7c). For deterministic forecast, Figure 7b shows that the RMSE for Meteo-Hydro streamflow forecast ranges from 250 to 350 m^3/s, while the RMSE for ESP-Hydro streamflow forecast ranges from 250 to 390 m^3/s. The Meteo-Hydro approach also performs better than ESP-Hydro with lower RMSE at all lead times especially after 3 days, with the average reduction of RMSE reaching 6% (Figure 7d).

Figure 7 also shows that both forecast skills have a similar diurnal cycle, where RMSE and CRPS reach their peaks around 00UTC and drop to their lows at 06UTC. Figure 8 shows the diurnal cycle of model employed variables, which are observed catchment mean rainfall, observed streamflow at Yantan and Longtan hydrological gauges, to explain the diurnal cycle of ESP-Hydro and Meteo-Hydro forecasting skills. These three input variables show different diurnal patterns. The observed rainfall starts to rise at 00UTC and reaches its maximum at 06UTC. The observed streamflow at Yantan hydrological gauge drops to its minimum at 12UTC and rises to its maximum at 00UTC. The streamflow from upstream Longtan hydrological gauge starts to drop at 00UTC and reaches its minimum at 06UTC. After comparing these diurnal cycles with the cycle of forecast skill, it is found that the forecast skill decreases when the upstream Longtan outflow starts to decrease, and the precipitation starts to increase. When the upstream Longtan outflow increases and the precipitation starts to decrease (after 06UTC), the forecast skill rises. Such information indicates that the hydrological model performs worse in the case of heavier rainfall event, and the decrease of upstream outflow may amplify such degradation when the portion of interval rainfall-runoff increased.

3.4 Meteo-Hydro-LSTM streamflow forecast

Machine learning methods can recognize patterns hidden in input data and can simulate or predict streamflow without explicit descriptions of the underlying physical processes. Figure 9 shows the RMSE of Meteo-Hydro-LSTM streamflow forecast using the ensemble mean hydrological forecast as described in the section above, and the past 6-hour observed streamflow of Yantan hydrological gauge as input. Compared with Meteo-Hydro and ESP-Hydro approach, applying LSTM model can further decrease the RMSE within the first 72 hours. The RMSE of Meteo-Hydro-LSTM approach ranges from 205 to 363 m^3/s during these three days, suggesting an average of 6% improvement against Meteo-Hydro approach.

Figure 9 also shows the RMSE of LSTM streamflow forecast only using the past 6-hour observed streamflow of Yantan hydrological gauge as input. Without using the physical model forecast, RMSE is improved only when the lead time is less than 1 day. And the performance of LSTM is far worse than Meteo-Hydro streamflow forecast when lead time is more than 2 days.

Figure 10 shows several examples of streamflow forecasts by Meteo-Hydro-LSTM approach and Meteo-Hydro approaches to show the forecast improvements in details. The Meteo-Hydro-LSTM approach reduced the flood peak value and the water loss during flood recession period compared with Meteo-Hydro streamflow forecast approach, which improves the streamflow prediction for most cases (Figs. 10b-10f). However, when the upstream reservoir's flood operation is triggered by continuous heavy rain, the Meteo-Hydro may underpredict the streamflow. With the LSTM model further decreases the streamflow, the Meteo-Hydro-LSTM method can end up with worsening the streamflow forecast, which means the machine learning method may improve forecasts when trained in different flood operating situations (Figure 10a).

4. Conclusions

In this study, we developed and evaluated a streamflow forecasting framework by coupling meteorological forecasts with a land surface hydrological model (CSSPv2) and a machine learning method (LSTM) over a cascade reservoir catchment using hindcast data from 2013 to 2017. The monthly observed runoff was used to calibrate the runoff generation module of the CSSPv2 model grid by grid, and the hourly observed streamflow at Yantan hydrological gauge was used to calibrate the routing module of the CSSPv2 model. Then, the bias-corrected TIGGE-ECMWF ensemble forecasts were used to drive the CSSPv2 for streamflow forecasts, and the LSTM model was used to correct the streamflow forecasts, resulted in an integrated meteorological-hydrological-machine learning forecast framework.

With automatic offline calibration of the CSSPv2 model, and the NSE values are 0.96, 0.92 and 0.61 for streamflow simulations at the Yantan gauge at monthly, daily and hourly time scales, respectively. The bias-corrected ensemble mean TIGGE-ECMWF forcings which perform the best among all ensemble members, show average RMSE and correlation of 14.6 mm/day and a 0.44 for precipitation forecasts, and 1.3 K and 0.87 for surface air temperature forecasts. By comparing with the hourly observed streamflow, the integrated hydrometeorological forecast approach

(Meteo-Hydro) increases the probabilistic and deterministic forecast skill against the initial condition-based approach (ESP-Hydro) by 6%.

Adding LSTM model to the hydrometeorological forecast (Meteo-Hydro-LSTM) can further reduce the forecast error. Within the first 72 hours, LSTM can improve the forecast skill with a maximum of 25% and an average of 6%. However, if we do not use the streamflow predicted by Meteo-Hydro, the error from the LSTM increases rapidly after 24 hours, and the historical data-based LSTM method performs worse than the Meteo-Hydro method. Most cascade reservoirs yet cannot forecast streamflow beyond 6 hours, and the integrated Meteo-Hydro-LSTM approach has potential to improve the forecasts at long leads. This study mainly focused on exploring the added values of meteorology-hydrology coupled forecast and LSTM forecast in a non-closed catchment, so the forecast uncertainty from upstream outflow was ignored by using the observed outflow. In the future, the upstream outflow forecast is planned to include, but this requires the development of upstream hydrometeorological forecast capability, as well as the reservoir regulation forecast that is very challenging. The artificial intelligence (AI) techniques are expected to complement the physical model for reservoir regulation forecast.

Acknowledgement. This work was supported by National Key R&D Program of China (2018YFA0606002), National Natural Science Foundation of China (41875105), and Natural Science Foundation of Jiangsu Province for Distinguished Young Scholars (BK20211540).

Data availability. The TIGGE-ECMWF hindcast data can be downloaded from https://apps.ecmwf.int/datasets/data/tigge/levtype=sfc/type=pf/ (Parsons et al., 2017), the in-situ observations and simulation data are available upon request.

References

Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., et al.: Two decades of anarchy? emerging themes and outstanding challenges for neural network river forecasting. Prog. Phys. Geogr. 36(4), 480-513. https://doi.org/10.1177/0309133312444943, 2012.

Adnan, R.M., et al.: Daily streamflow prediction using optimally pruned extreme learning machine. J. Hydrol. 577. https://doi.org/10.1016/j.jhydrol.2019.123981, 2019.

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., & Pappenberger, F.: GloFAS-global ensemble streamflow forecasting and flood early warning. Hydrol. Earth Syst. Sci. 17 (3), 1161–1175. https://doi.org/10.5194/hess17-1161-2013, 2013.

Balint, G., Csik, A., Bartha, P., Gauzer, B., Bonta, I.: Application of meterological ensembles for Danube flood forecasting and warning. In: Marsalek, J., Stancalie, G., Balint, G. (Eds.), Transboundary Floods: Reducing Risks through Flood Management. Springer, NATO Science Series, Dordecht, The Netherlands, pp. 57–68. https://doi.org/10.1007/1-4020-4902-1_6 , 2006.

Bougeault, P., et al.: The THORPEX interactive grand global ensemble, Bull. Am. Meteorol. Soc., 91, 1059–1072. http://dx.doi.org/10.1175/2010BAMS2853.1, 2010.

Dai, Y., Dickinson, R. E., Wang, Y. P.: A two-big-leaf model for canopy temperature, photosynthesis, and stomatal conductance. J. Clim. 17(12),2281–2299. https://doi.org/10.1175/1520-0442(2004)017<2281:ATMFCT>2.0.CO;2 , 2004.

Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., et al.: The Common Land Model. Bull. Am. Meteorol. Soc. 84(8), 1013–1024. https://doi.org/10.1175/BAMS-84-8-1013, 2003.

Dang, T. D., Chowdhury, A. K., Galelli, S.: On the representation of water reservoir storage and operations in large-scale hydrological models: implications on model parameterization and climate change impact assessments. Hydrol. Earth Syst. Sci., 24, 397–416. https://doi.org/10.5194/hess-24-397-2020, 2020.

Day, G.N.: Extended Streamflow Forecasting Using NWSRFS. J. Water Resour. Plan Manag. 111 (2): 157-170, 1985.

Duan, Q., Sorooshian, S., Gupta, V. K.: Optimal use of SCEUA global optimization method for calibrating watershed models, J. Hydrol., 158, 265–284, https://doi.org/10.1016/0022-1694(94)90057-4, 1994.

Gao, X., Zeng, Y., Wang, J., Liu, H.: Immediate impacts of the second impoundment on fish communities in the Three Gorges Reservoir, Environ. Biol. Fish., 87, 163–173. https://doi.org/10.1007/s10641-009-9577-1., 2010.

Getirana, A. C. V., Boone, A., Yamazaki, D., Decharme, B., Papa, F., and Mognard, N.: The Hydrological Modeling and Analysis Platform (HyMAP): Evaluation in the Amazon Basin, J. Hydrometeorol., 13, 1641–1665, https://doi.org/10.1175/JHM-D-12-021.1, 2012.

Graaf, I. D., Sutanudjaja, E. H., Beek, L. V., et al.: A high-resolution global-scale groundwater model. Hydrol. Earth Syst. Sci. 19(2):823-837. https://doi.org/10.5194/hess-19-823-2015, 2015.

Hao, Z., Aghakouchak, A., Phillips, T. J.: Changes in concurrent monthly precipitation and temperature extremes. Environ. Res. Lett. 8(3), 1402-1416. https://doi.org/10.1088/1748-9326/8/3/034014, 2013.

Hopson, T., Webster, P.: A 1–10 day ensemble forecasting scheme for the major river basins of Bangladesh: forecasting severe floods of 2003–2007. J. Hydrometeorol. 11(3), 618-641. https://doi.org/10.1175/2009JHM1006.1, 2010.

Hornberger, G. M., E. W. Boyer.: Recent advances in watershed modeling, U.S. Natl. Rep. Int. Union Geod. Geophys. 1991 – 1994, Rev. Geophys., 33, 949 – 957. https://doi.org/10.1029/95RG00288, 1995.

Humphrey, G.B., Gibbs, M.S., Dandy, G.C., Maier, H.R.: A hybrid approach to monthly streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial neural network. J. Hydrol. 540, 623–640. https://doi.org/10.1016/j.jhydrol.2016.06.026, 2016.

Jasper, K., Gurtz, J., Lang, H.: Advanced flood forecasting in Alpine watersheds by coupling meteorological observations and forecasts with a distributed hydrological model. J. Hydrol. 267 (1–2), 40–52. https://doi.org/10.1016/S0022-1694(02)00138-5, 2002.

Jaun, S., Ahrens, B., Walser, A., Ewen, T., Schär, C.: A probabilistic view on theAugust 2005 floods in the upper Rhine catchment. Nat. Hazard Earth Sys. 8, 281–291. https://doi.org/10.5194/nhess-8-281-2008, 2008.

Ji, P., X. Yuan., Y. Jiao., C. Wang., S. Han., C. Shi.: Anthropogenic contributions to the 2018 extreme flooding over the upper Yellow River basin in China. Bull. Am.

Meteorol. Soc. 101(1), S89-S94, https://doi.org/10.1175/BAMS-D-19-0105.1, 2020.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resour. Res. 42, 1–5. https://doi.org/10.1029/2005WR004362, 2006.

Kisi, O.: Streamflow forecasting using different artificial neural network algorithms. J. Hydrol. Eng. 12 (5), 532–539. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:5(532), 2007.

Kollet, S. J., Maxwell, R. M., Woodward, C. S., Smith, S., Vanderborght, J., & Vereecken, H., et al.: Proof of concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources. Water Resour. Res. 46(4). https://doi.org/10.1029/2009WR008730, 2010.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M.: Rainfall–runoff modelling using long short-term memory (LSTM) networks. Hydrol. Earth Syst. Sci. 22 (11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018, 2018.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., Nearing, G. S.: Towards Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning. Water Resour. Res. 55. https://doi.org/10.1029/2019wr026065, 2019.

Leith, C. E.: Theoretical skill of monte carlo forecasts. Mon. Weather Rev. 102(6), 409-418. https://doi.org/10.1175/1520-0493(1974)1022.0.CO;2, 1974.

Luo, X., Li, H. Y., Ruby, L. L., Tesfa, T. K., Augusto, G., & Fabrice, P., et al.:

Modeling surface water dynamics in the amazon basin using mosart-inundation v1.0: impacts of geomorphological parameters and river flow representation. Geosci. Model. Dev., 10(3), 1-42. https://doi.org/10.5194/gmd-10-1233-2017 , 2017.

Mulvaney, T. J.: On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and of flood discharges in a given catchment, in: Proceedings Institution of Civil Engineers, Dublin, Vol. 4, 18–31, 1850.

Pappenberger, F., Ramos, M. H., Cloke, H. L., Wetterhall, F., Alfieri, L., Bogner, K., et al.: How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction. J. Hydrol. 522, 697–713. https://doi.org/10.1016/j.jhydrol.2015.01.024, 2015.

Parsons, D. B., Beland, M., Burridge, D., Bougeault, P., Brunet, G., Caughey, J., et al.: Thorpex research and the science of prediction. Bull. Am. Meteorol. Soc., 98, 807-830, https://doi.org/10.1175/BAMS-D-14-00025.1 , 2017.

Robertson, D. E., Wang, Q. J.: Seasonal Forecasts of Unregulated Inflows into the Murray River, Australia. Water. Resour. Manag. 27(8):2747–2769. https://doi.org/10.1007/s11269-013-0313-4, 2013.

Shao, J., Wang, J., Lv, S., Bing, J.: Spatial and temporal variability of seasonal precipitation in Poyang Lake basin and possible links with climate indices. Hydrol. Res. 47(S1):51–68. https://doi.org/10.2166/nh.2016.249, 2016.

Toth, Z., Zhu, Y., Marchok, T.: The use of ensembles to identify forecasts with small and large uncertainty. Weather Forecast 16(4), 463-463. https://doi.org/10.1175/1520-0434(2001)0162.0.CO;2, 2001.

Wang, R., Zhang, J., Guo, E., Zhao, C., Cao, T.: Spatial and temporal variations of precipitation concentration and their relationships with large-scale atmospheric circulations across Northeast China. Atmos. Res. 222:62–73. https://doi.org/10.1016/j.atmosres.2019.02.008, 2019.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Volume 91, Second Edition (International Geophysics), 2005.

Wood, E. F., et al.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. Water Resour. Res., 47, W05301, https://doi.org/10.1029/2010WR010090, 2011.

Xu, Y.P., Gao, X., Zhang, Y., Kang, L.: Coupling a regional climate model and distributed hydrological model to assess future water resources in Jinhua River Basin, East China. ASCE. J. Hydrol. Eng. 20, 2015, https://doi.org/10.1061/(ASCE) HE.1943-5584.0001007, 2015.

Yang, S., Yang, D., Chen, J., Santisirisomboon, J., Zhao, B.: A physical process and machine learning combined hydrological model for daily streamflow simulations of large watersheds with limited observation data. J. Hydrol. 590, 125206. https://doi.org/10.1016/j.jhydrol.2020.125206, 2020.

Yaseen, Z.M., Sulaiman, S.O., Deo, R.C., Chau, K.-W.: An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. J. Hydrol. 569, 387–408. https://doi.org/10.1016/j.jhydrol.2018.11.069, 2018.

Ye, A., Duan, Q., Yuan, X., Wood, E. F., Schaake, J.: Hydrologic post-processing of MOPEX streamflow simulations. J. Hydrol. 508, 147-156, doi:10.1016/j.jhydrol.2013.10.055, 2014

Yuan, X., Ma, F., Wang, L., Zheng, Z., Ma, Z., Ye A., Peng, S.: An experimental seasonal hydrological forecasting system over the Yellow River basin-Part 1: Understanding the role of initial hydrological conditions. Hydrol. Earth Syst. Sci. 20, 2437–2451, https://doi.org/10.5194/hess-20-2437-2016, 2016.

Yuan, X., S. Wang, and Z.-Z. Hu, 2018a: Do climate change and El Niño increase likelihood of Yangtze River extreme rainfall? Bull. Am. Meteorol. Soc. 99, S113-S117, https://doi.org/10.1175/BAMS-D-17-0089.1, 2018a.

Yuan, X., Ji, P., Wang, L., Liang, X. Z., Yang, K., Ye, A., et al.: High‐resolution land surface modeling of hydrological changes over the sanjiangyuan region in the eastern tibetan plateau: 1. model development and evaluation. J. Adv. Model Earth Syst. https://doi.org/10.1029/2018MS001412, 2018b.

Zhang, Y., Erkyihum, S. T., Block, P.: Filling the GERD: evaluating hydroclimatic variability and impoundment strategies for Blue Nile riparian countries, Water Int., 41, 593–610. https://doi.org/10.1080/02508060.2016.1178467, 2016.

Zhao, T.T.G., Cai, X.M., Yang, D.W.: Effect of streamflow forecast uncertainty on real-time reservoir operation. Adv. Water Resour. 34 (4), 495–504, https://doi.org/10.1016/j.advwatres.2011.01.004, 2011.

Zhu, E., X. Yuan., A. Wood.: Benchmark Decadal Forecast Skill for Terrestrial Water

Storage Estimated by an Elasticity Framework. Nat. Commun. 10, 1237, https://doi.org/10.1038/s41467-019-09245-3, 2019.

Table 1. Information of hydrological gauges.

Gauge	Longitude (°E)	Latitude (°N)	Drainage area (km^2)
Longtan	107.09	25.00	-
Yantan	107.50	24.11	5950 (orange area in Fig. 1)
Luofu	107.36	24.90	800 (green area in Fig. 1)
Jiazhuan	107.12	24.21	2150 (purple area in Fig. 1)

Table 2. Information of hydrological datasets

Dataset	Time Range	Time step
Rain Gauge Observation Forcing	2013/1/1 ~ 2017/12/31	Hourly
Longtan & Yantan Discharge Gauge Streamflow data	2013/1/1 ~ 2017/12/31	Hourly
Jiazhuan & Luofu Discharge Gauge Streamflow data	2013/4/1 ~ 2017/9/30	Daily
TIGGE-ECMWF Forecast Forcing	2013/4/1 ~ 2017/9/30	Hourly

Table 3. Descriptions of calibrated parameters

Parameters	Range
Maximum velocity of baseflow (mm/day)	0.00000116 ~ 0.000579
Fraction of maximum velocity of baseflow where non-linear baseflow begins	0.001 ~ 0.99
Fraction of maximum soil moisture where non-linear baseflow occurs	0.2 ~ 0.99
Variable infiltration curve parameter	0.001 ~ 1
River width (m)	0 ~ 101.16
River depth (m)	0 ~ 6.46
River density (km/km^2)	0.049 ~ 1.03
River roughness	0.033 ~ 0.05
River slope	0.015 ~ 0.47

Table 4. Experimental design in this study.

Experiments	Description
ESP-Hydro	Using CSSPv2 land surface hydrological model driven by randomly-sampled historical meteorological forcings
Meteo-Hydro	Using CSSPv2 model driven by bias-corrected TIGGE-ECMWF hindcast meteorological forcings
Meteo-Hydro-LSTM	Using LSTM model to correct streamflow from Meteo-Hydro hindcast
LSTM	Using LSTM model to forecast streamflow based on observation only

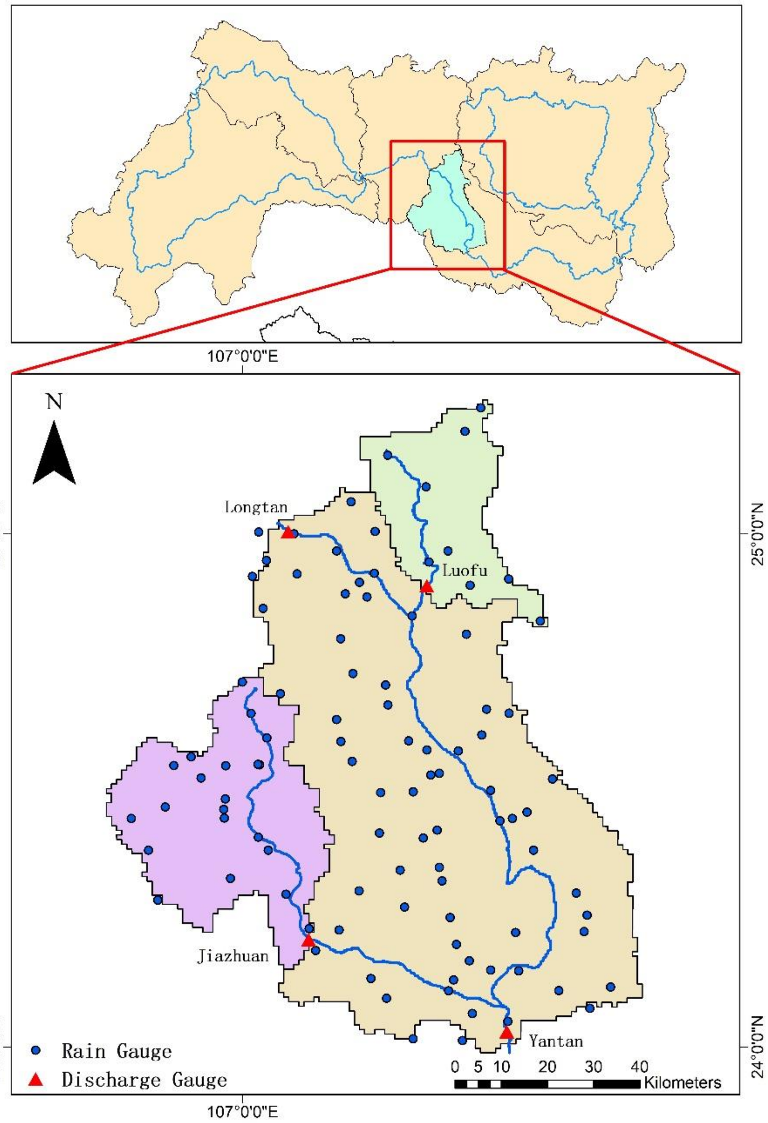

Figure 1. Locations of discharge gauges and rain gauges over the Yantan basin.

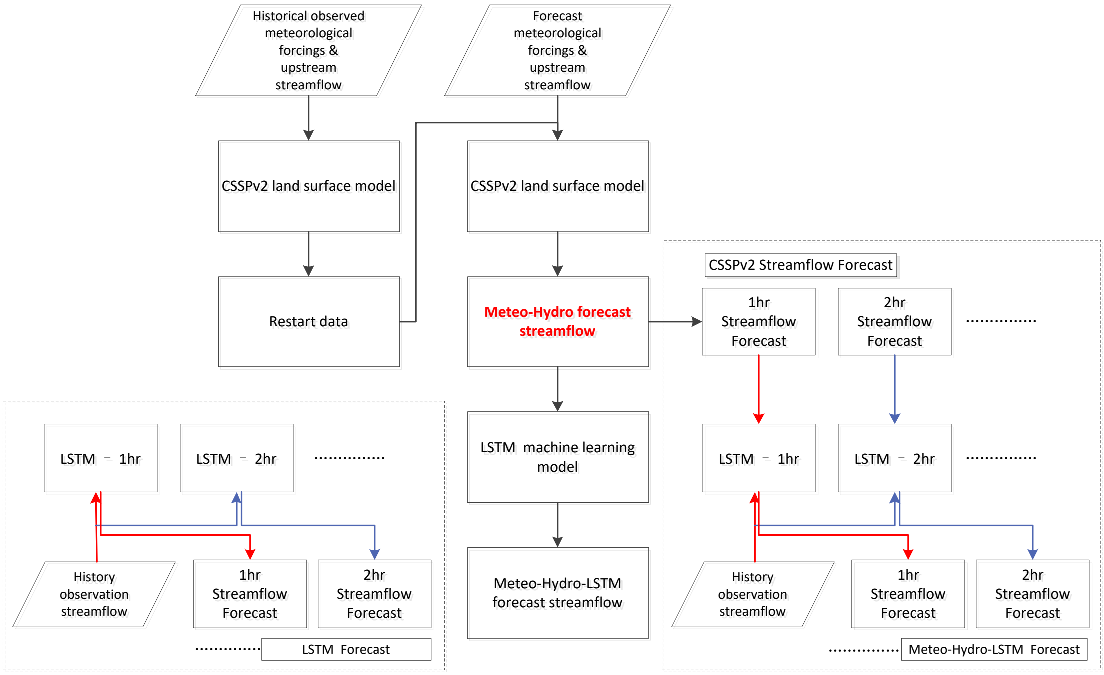

Figure 2. A diagram for the integrated hydrometeorological and machine learning streamflow prediction.

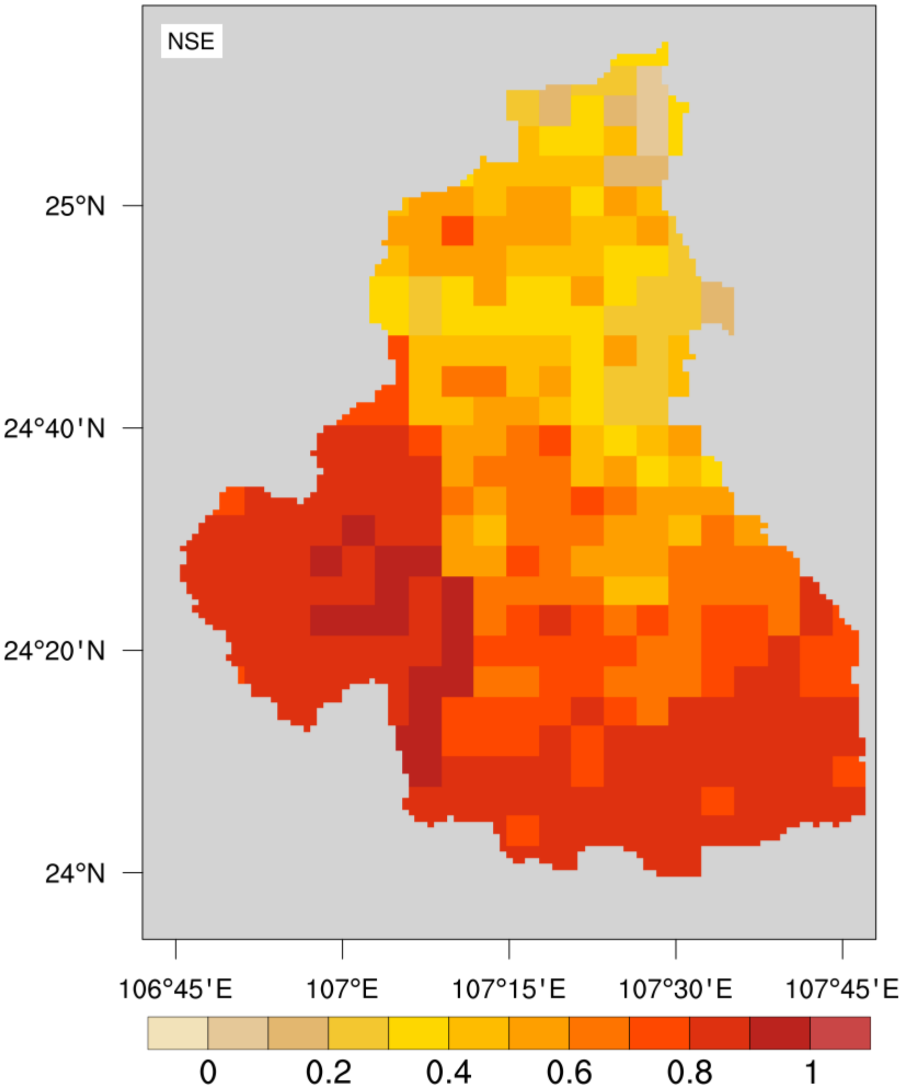

Figure 3. Nash-Sutcliff efficiency coefficients for the calibrated grid runoff simulation from CSSPv2.

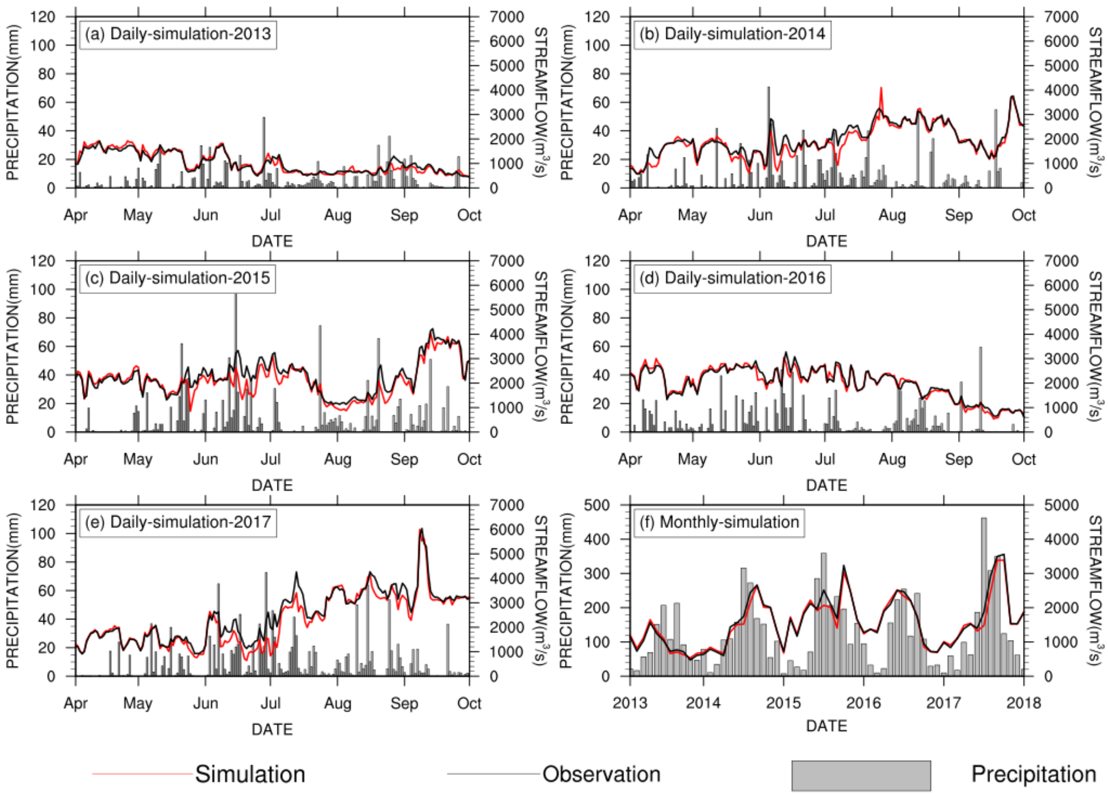

Figure 4. Evaluation of streamflow simulations at Yantan gauge. The black and red lines are observed and simulated streamflow. (a)-(e) are for daily streamflow, and (f) is for monthly streamflow. The gray bars represent daily (or monthly) precipitation.

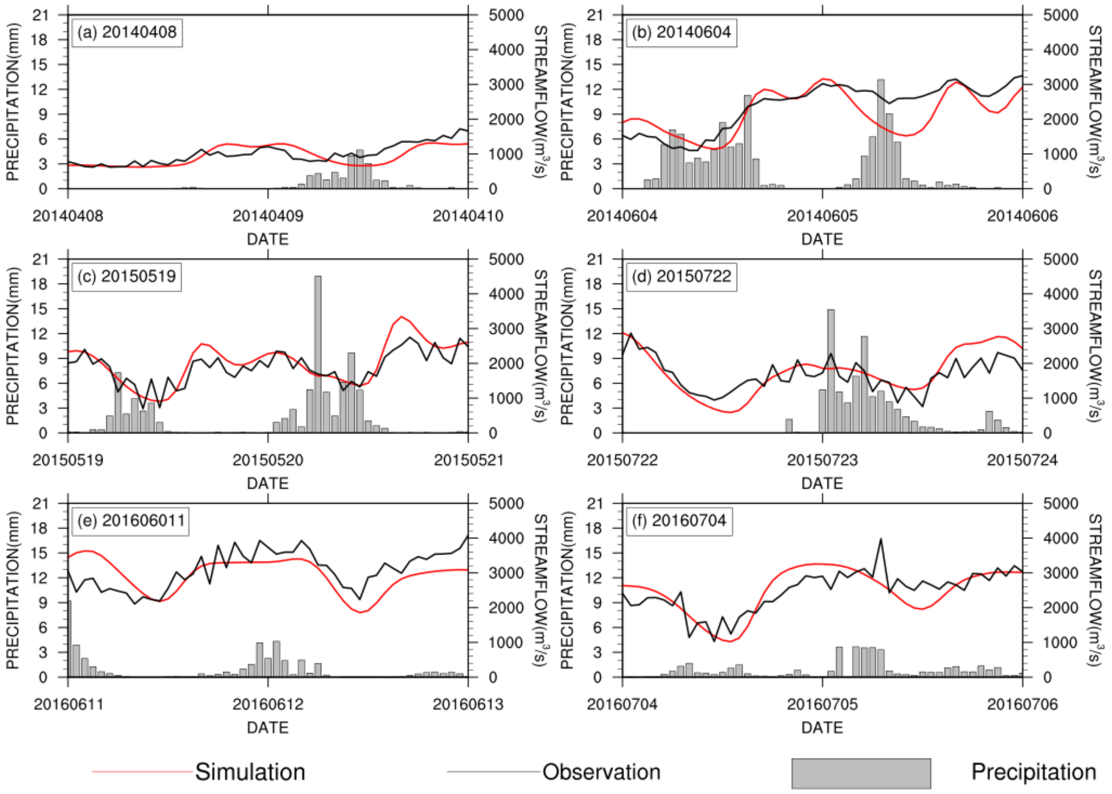

Figure 5. The same as Figure 4, but for the evaluation of hourly streamflow simulations at Yantan gauge.

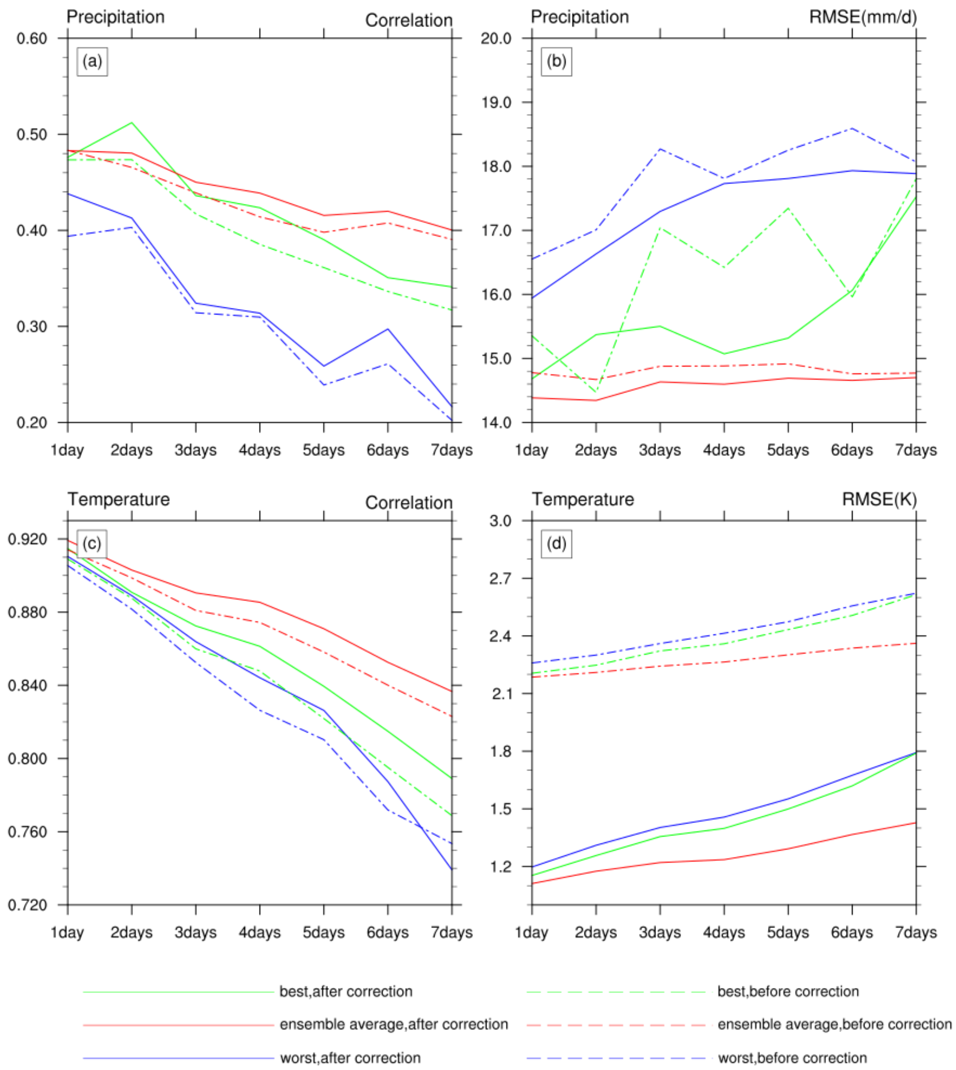

Figure 6. Evaluation of precipitation and temperature hindcasts from TIGGE-ECMWF. The red and blue lines represent the best and worst results among 51 TIGGE-ECMWF ensemble members respectively, and the green lines represent the results for the ensemble means of 51 members. Solid and dashed lines represent the results after and before bias corrections, respectively.

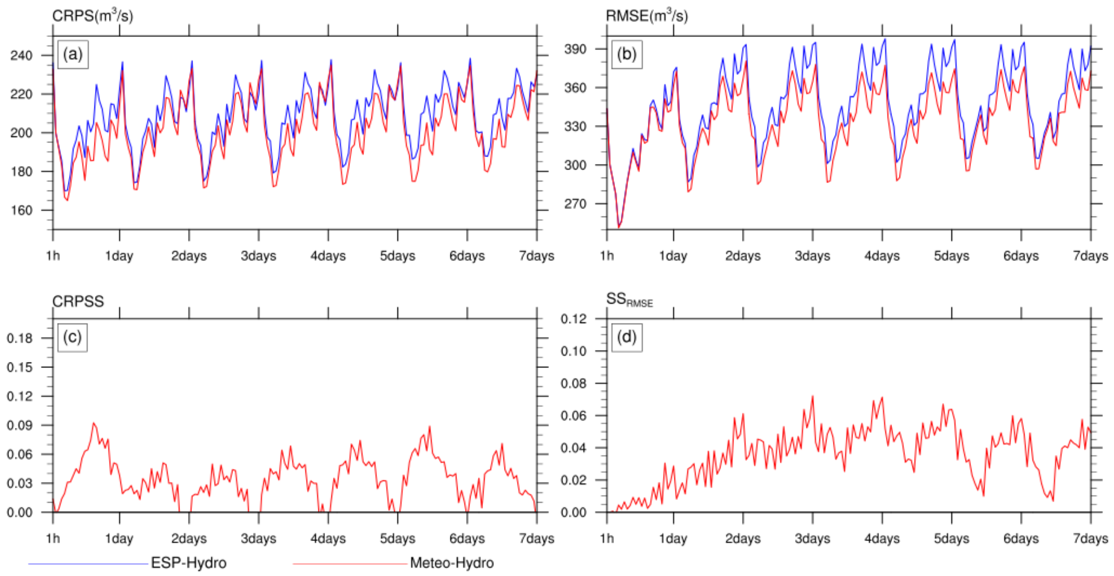

Figure 7. (a) Continuous Ranked Probability Score (CRPS) and (b) Root Mean Squared Error (RMSE) for daily streamflow ensemble forecasts at Yantan gauge. (c) and (d) are the skill score in terms of CRPS and RMSE for Meteo+Hydro, where ESP+Hydro is used as reference forecast.

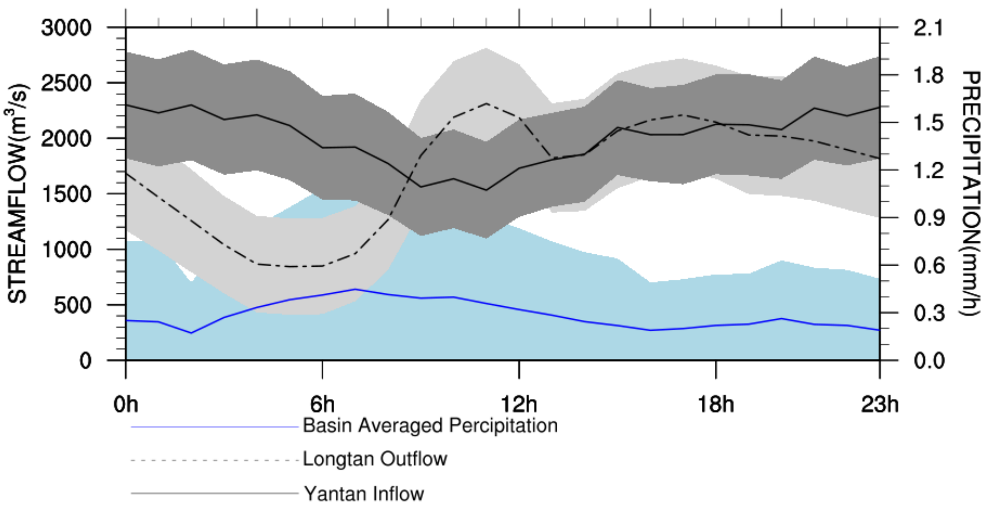

Figure 8. Diurnal cycle of Longtan outflow (m^3/s; dashed black line), Yantan inflow (m^3/s; solid black line) and basin-averaged precipitation (mm/h; blue line).

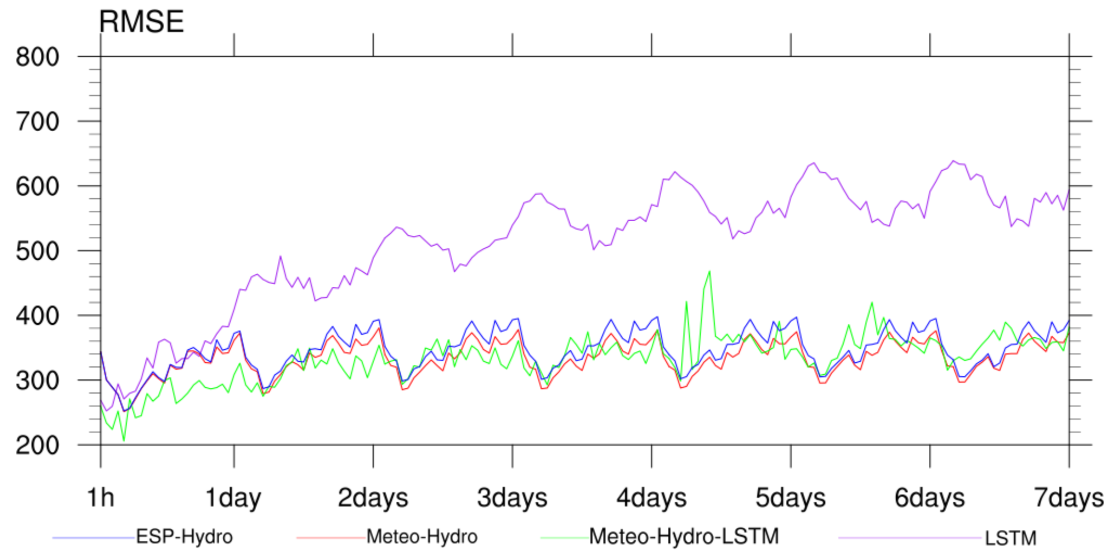

Figure 9. RMSE (m^3/s) for hourly streamflow hindcasts from four forecast approaches. The green line represents the Meteo+Hydro+LSTM forecast, the red line represents the Meteo+Hydro forecast, the blue line represent the ESP+Hydro forecast, and the purple line represents the LSTM forecast based on historical streamflow observation alone.

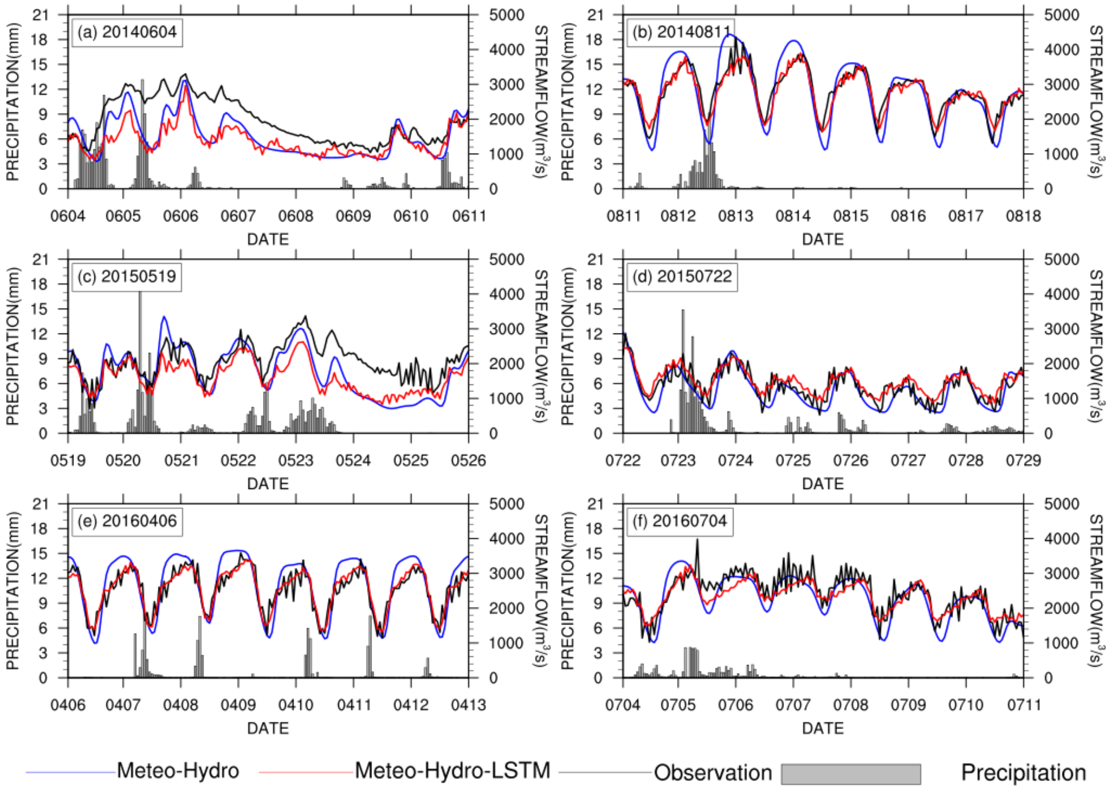

Figure 10. Evaluation of the forecast approaches for a few flooding events. The black lines are observed streamflow from Yantan hydrological gauge, the blue lines are the Meteo+Hydro ensemble mean streamflow forecast, and the red lines are the Meteo+Hydro+LSTM forecast streamflow by using Meteo+Hydro ensemble mean forecast with LSTM. The gray bars represent hourly precipitation averaged over the basin.